



Defense Threat Reduction Agency
8725 John J. Kingman Road, MS
6201 Fort Belvoir, VA 22060-6201



DTRA-TR-12-63

TECHNICAL REPORT

Collaborative Research: Catalog Completeness and Accuracy

Approved for public release, distribution is unlimited.

January 2013

DTRA01-00-C-0062

Frank L.Vernon

Prepared by:
University of California – San
Diego
Scripps Institution of
Oceanography
9500 Gilman Drive
La Jolla, CA 92093

DESTRUCTION NOTICE:

Destroy this report when it is no longer needed.
Do not return to sender.

PLEASE NOTIFY THE DEFENSE THREAT REDUCTION
AGENCY, ATTN: DTRIAC/ J-3 ONIUI , 8725 JOHN J. KINGMAN ROAD,
MS-6201, FT BELVOIR, VA 22060-6201, IF YOUR ADDRESS
IS INCORRECT, IF YOU WISH THAT IT BE DELETED FROM THE
DISTRIBUTION LIST, OR IF THE ADDRESSEE IS NO
LONGER EMPLOYED BY YOUR ORGANIZATION.

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 00-01-2013			2. REPORT TYPE Technical		3. DATES COVERED (From - To) 30 June 2000 - 31 Dec 2004	
4. TITLE AND SUBTITLE Collaborative Research: Catalog Completeness and Accuracy					5a. CONTRACT NUMBER DTRA01-00-C-0062	5b. GRANT NUMBER
					5c. PROGRAM ELEMENT NUMBER 463D	
6. AUTHOR(S) Frank L. Vernon					5d. PROJECT NUMBER CD	5e. TASK NUMBER CD
					5f. WORK UNIT NUMBER PRDAO	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California - San Diego Scripps Institution of Oceanography 9500 Gilman Drive La Jolla, CA 92093					8. PERFORMING ORGANIZATION REPORT NUMBER DTRA20-1080	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Threat Reduction Agency 8725 John J. Kingman Road, MS 6201 Fort Belvoir, VA 22060-6201 PM/August Taylor					10. SPONSOR/MONITOR'S ACRONYM(S) DTRA	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S) DTRA-TR-12-63	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.						
13. SUPPLEMENTARY NOTES This work was sponsored by the Defense Threat Reduction Agency under RDT&E RMC Code 340.						
14. ABSTRACT We have assembled data from a number of portable seismic experiments mounted in the last few years driven largely by an interest in the structure and dynamics of the India-Asia collision. We have used data collected in these experiments, as well as broadband data collected on a number of national and private networks, along with global seismic stations to produce a high quality regional catalogue with a detection threshold less than magnitude 3 for most of the middle east and central Asia. We developed a new grid-based implementation of the progress multiple event location. A second parallel effort is to apply waveform correlation methods to the entire dataset. Waveform correlation can dramatically improve measurement precision, but it can only be done successfully when waveforms are similar enough to allow correlation.						
15. SUBJECT TERMS Earthquakes, Location Algorithms, Central Asia						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 27	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)	

CONVERSION TABLE

Conversion Factors for U.S. Customary to metric (SI) units of measurement.

MULTIPLY → BY ← DIVIDE
TO GET ← BY → TO GET

angstrom	1.000 000 x E -10	meters (m)
atmosphere (normal)	1.013 25 x E +2	kilo pascal (kPa)
bar	1.000 000 x E +2	kilo pascal (kPa)
barn	1.000 000 x E -28	meter ² (m ²)
British thermal unit (thermochemical)	1.054 350 x E +3	joule (J)
calorie (thermochemical)	4.184 000	joule (J)
cal (thermochemical/cm ²)	4.184 000 x E -2	mega joule/m ² (MJ/m ²)
curie	3.700 000 x E +1	*giga bacquerel (GBq)
degree (angle)	1.745 329 x E -2	radian (rad)
degree Fahrenheit	t _k = (t°f + 459.67)/1.8	degree kelvin (K)
electron volt	1.602 19 x E -19	joule (J)
erg	1.000 000 x E -7	joule (J)
erg/second	1.000 000 x E -7	watt (W)
foot	3.048 000 x E -1	meter (m)
foot-pound-force	1.355 818	joule (J)
gallon (U.S. liquid)	3.785 412 x E -3	meter ³ (m ³)
inch	2.540 000 x E -2	meter (m)
jerk	1.000 000 x E +9	joule (J)
joule/kilogram (J/kg) radiation dose absorbed	1.000 000	Gray (Gy)
kilotons	4.183	terajoules
kip (1000 lbf)	4.448 222 x E +3	newton (N)
kip/inch ² (ksi)	6.894 757 x E +3	kilo pascal (kPa)
ktap	1.000 000 x E +2	newton-second/m ² (N-s/m ²)
micron	1.000 000 x E -6	meter (m)
mil	2.540 000 x E -5	meter (m)
mile (international)	1.609 344 x E +3	meter (m)
ounce	2.834 952 x E -2	kilogram (kg)
pound-force (lbs avoirdupois)	4.448 222	newton (N)
pound-force inch	1.129 848 x E -1	newton-meter (N-m)
pound-force/inch	1.751 268 x E +2	newton/meter (N/m)
pound-force/foot ²	4.788 026 x E -2	kilo pascal (kPa)
pound-force/inch ² (psi)	6.894 757	kilo pascal (kPa)
pound-mass (lbf avoirdupois)	4.535 924 x E -1	kilogram (kg)
pound-mass-foot ² (moment of inertia)	4.214 011 x E -2	kilogram-meter ² (kg-m ²)
pound-mass/foot ³	1.601 846 x E +1	kilogram-meter ³ (kg/m ³)
rad (radiation dose absorbed)	1.000 000 x E -2	**Gray (Gy)
roentgen	2.579 760 x E -4	coulomb/kilogram (C/kg)
shake	1.000 000 x E -8	second (s)
slug	1.459 390 x E +1	kilogram (kg)
torr (mm Hg, 0° C)	1.333 22 x E -1	kilo pascal (kPa)

*The bacquerel (Bq) is the SI unit of radioactivity; 1 Bq = 1 event/s.

**The Gray (GY) is the SI unit of absorbed radiation.

Abstract

We have assembled data from a number of portable seismic experiments mounted in the last few years drive largely by an interest in the structure and dynamics of the India-Asia collision. We have used data collected in these experiments, as well broadband data collected on a number of national and private networks, along with global seismic stations to produce a high quality regional catalogue with a detection threshold less than magnitude 3 for most of the middle East and central Asia. We developed a new grid-based implementation of the progressive multiple event location. A second parallel effort is to apply waveform correlation methods to the entire dataset. Waveform correlation can dramatically improve measurement precision, but it can only be done successfully when waveforms are similar enough to allow correlation.

Table of Contents

Abstract	ii
List of Tables	iv
List of Figures	iv
Summary	1
Introduction	2
Methods, Assumptions, and Procedures	2
Regional Phase Fundamentals	2
Results and Discussions	5
Catalog Preparation	5
Improving location accuracy and error appraisal	6
Conclusions	19
References	19

List of Tables

Table 1: Data Source for Catalogue Construction Summary	5
Table 2: Software Development	6

List of Figures

Figure 1: Sample regional seismogram.	4
Figure 2: Station map showing all network and portable broadband stations.	7
Figure 3: Seismicity recorded by regional networks and portable experiments.	8
Figure 4: Example of travel time surface defined with a GCLgrid object.	10
Figure 5: Simulation results from grid based PMEL implementation.	13
Figure 6: Example of grid-based PMEL processing on KNET.	15
Figure 7: Example of prototype cross-correlation procedure.	17
Figure 8: Effects of cross-correlation and PMEL on aftershock locations	18

Collaborative Research: Seismic Catalogue Completeness and Accuracy

Summary

We have assembled data from a number of portable seismic experiments mounted in the last few years drive largely by an interest in the structure and dynamics of the India-Asia collision. We have used data collected in these experiments, as well broadband data collected on a number of national and private networks, along with global seismic stations to produce a high quality regional catalogue with a detection threshold less than magnitude 3 for most of the middle East and central Asia. Each network or experiment listed has been individually processed and has its own catalog. Summing the catalogs from each network or experiment provides over 40,000 events. We developed a new grid-based implementation of the progressive multiple event location (PMEL). We use a spatial association algorithm to associate each events with one or more grid points within a region. We then apply PMEL to each spatial grouping (cluster) to estimate two quantities: (1) revised estimates of the hypocenters of every event in each ensemble, and (2) estimates of path corrections for each station and each seismic phase. We produce these in hierarchy of scales. We start at the smallest scale to relocate events in the vicinity of each of the individual networks. We use distance weighting to dampen the influence on distant stations that would otherwise skew locations of the largest events that light up a larger area relative to smaller events that are only recorded at the closest stations. We then use the local grid travel times in combination with the “ttregions” travel-time calculator to build a control framework for a regional solution on a coarser grid spanning the Middle-East and most of southern Asia. Events falling inside the local-scale grids will automatically use the local grid corrections as the 3D reference model to derive an improved absolute location framework. This will link the local scale network results into a larger scale framework.

A second parallel effort is to apply waveform correlation methods to the entire dataset. Waveform correlation can dramatically improve measurement precision, but it can only be done successfully when waveforms are similar enough to allow correlation. A key research question is the distance scale length over which waveforms can be correlated. A key practical problem is mixing hand-picked and cross-correlation measurements in the same framework. The working catalogue we produced is the starting point for several critical research questions important for nuclear monitoring.

Collaborative Research: Seismic Catalogue Completeness and Accuracy

Introduction

Seismic event detection and location are the single most important research issues for adequately monitoring underground testing of nuclear weapons. Confidence in seismic monitoring relies on the assumption that any underground nuclear weapons test will be detected at an adequate number of seismic stations to allow it to be located with sufficient accuracy that the test can be confirmed by on-site inspection. Another way to put this is that confidence in seismically monitoring is equivalent to confidence in the *seismic catalogue* that is produced by the monitoring system. There are two principle components of a catalogue: (1) completeness in terms of representing all the seismicity within the region of monitoring interest (referred to here as *detection*), and (2) the accuracy of the source parameters for the events within the catalogue. The accuracy of the locations in the catalogue is an essential component of calibration; the methods used to construct accurate locations must be transportable to the DoE knowledge database.

We have focused on constructing a catalogue for the region stretching from Saudi Arabia to western China for 1995 to the present. We have used all available data sources which includes several temporary, portable seismic experiments and private or national seismic networks that are not easily available to other research groups. The catalogue construction has two principle tasks: detecting and associating seismic phase and locating these events and assigning realistic location errors. The former is a large data processing effort that has required significant numbers of analyst time to pick seismic phases and merge databases from other sources. The catalog we have assembled under exceeds 40,000 events. This includes the unique Tien Shan experiment dataset which provides over 20,000 events with a magnitude completeness to $M = 2.0$ over a 10 degree by 10 degree region in central Asia. The second is a combined applied and basic research problem. It is “applied” because most of our efforts have been expended in implementing theoretical concepts worked out years ago into a workable computing framework. It has also been a “basic research” effort, however, because in the process of trying to improve location capabilities we have developed some new concepts in location methodology that we would argue have significant promise.

Methods, Assumptions, and Procedures

Regional Phase Fundamentals

The location of seismic events is always subject to a series of uncertainties, which affects both the accuracy, and precision of a reported hypocenter. Figure 1 illustrates two of

the fundamental problems we face.

1. *Earth model errors.* This is reflected here in two ways. The first is the most obvious but has a more subtle effect on location estimates. Notice that these data span the crossover between Pg and Pn. The analyst picks follow the data correctly, but Pn in this region (Tien Shan) arrives significantly later than that predicted by the model (iaspei91). The reason is that this event was located with an earth model with a thicker crust than iaspei91 which leads to a later predicted Pn time that is much more consistent with the data. This illustrates a critical principle that is too often misunderstood: travel time residuals from unconstrained sources mean nothing outside the context of the computer program that estimated the hypocenter. The location estimate and the residuals are inseparably coupled. The second element of this issue that Figure 1 addresses is more subtle. Notice the relatively large scatter in measured Pn arrival times. That is, the data as a whole line up on the Pn phase velocity (used to align the seismograms), but there is a scatter of more than 1 s around an imaginary vertical line through the Pn picks. This is a reflection of earth model errors caused by laterally varying structure, which in this case is undoubtedly due mainly to crustal thickness variations in the Tien Shan region.
2. *Phase identification and measurement precision.* The most common procedure for location is to make use of first arriving P waves. However, at regional distances this can be problematic as the data in Figure 1 illustrates. Pn is often very small or a series of multiple arrivals, and can not be picked and identified with confidence. Furthermore, as the size of event drops it is easy to make a very large blunder. Imagine Figure 1 with the noise level at all stations elevated by only a factor of 2 or 3. Pn would become invisible for most of these stations and the apparent first arrival would be Pg. The use of secondary phases (Pg, P*, Sn, Sg or Lg) in a location procedures has obvious benefits, but as this example illustrates, correctly identifying and timing these phases is extremely difficult, especially in automated systems. Bergman and Engdahl (2002) have developed a statistical framework for evaluating teleseismic secondary arrives, but at regional distances a more empirical approach is needed. The key message is that regional phases association is a serious problem and the timing precision of what we measure is normally orders of magnitude worse than impulsive arrivals from local events or teleseismic P waves.

A final issue about regional event location is outside the scope of Figure 1. It is what people in GPS processing call “constellation problems”. At teleseismic distances, large-sized seismic events can be located with high confidence using only first-arriving P phases if there are large numbers of recording stations that are well distributed in azimuth and distance. However, as a seismic event decreases in size, the number of recording stations decreases.

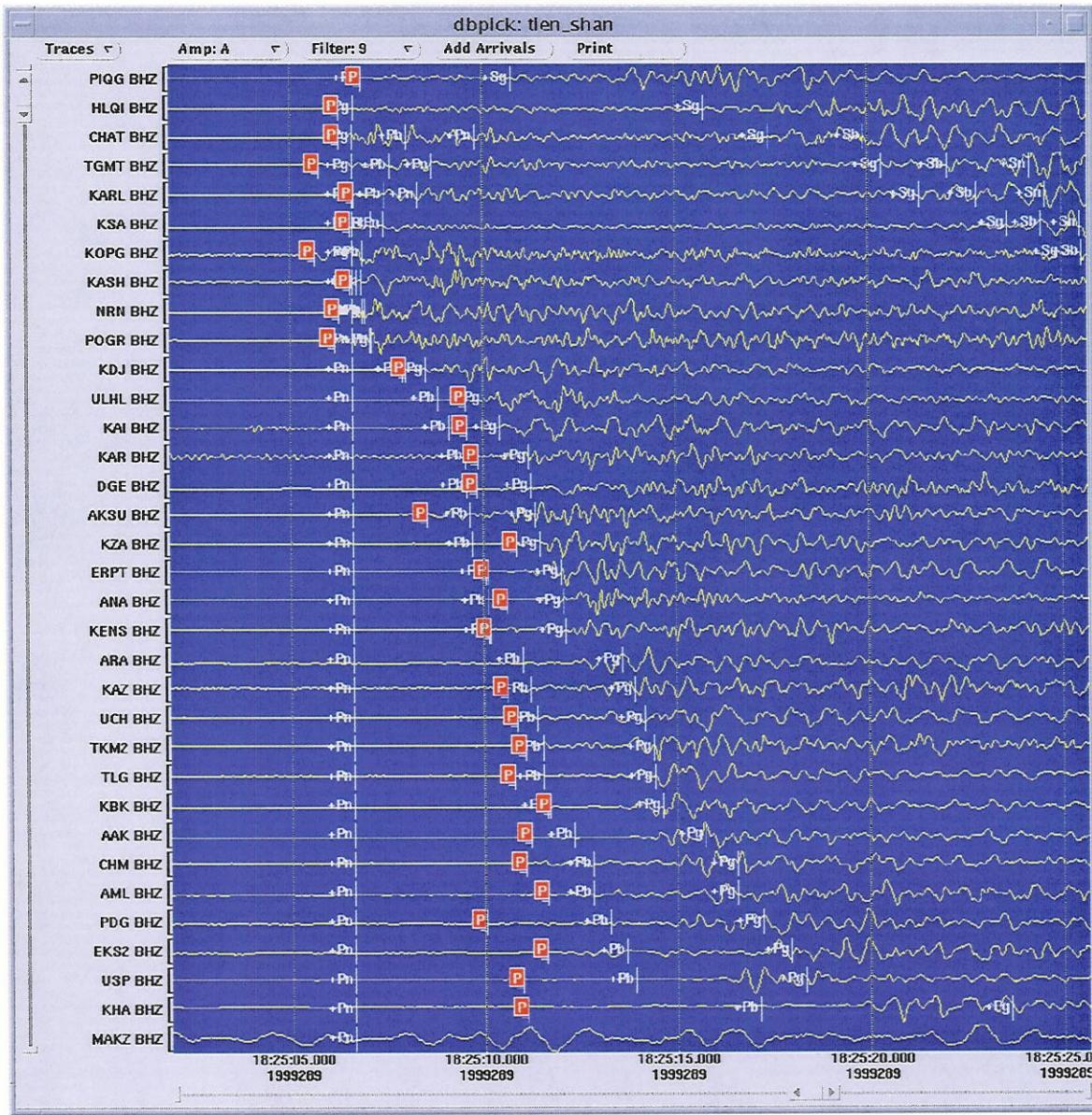


Figure 1. Sample regional seismogram showing fundamental problems in analysis of regional phases. The event shown is a magnitude 3.4 event located in the south-central Tien Shan. In this region the crustal thickness approaches 70 km. The seismograms have been aligned on the predicted first arrival time from IASPEI91 and arranged in order of increasing distance with the top seismogram being the closest station. Note the large delay in the measured first arrival time relative to that predicted for Pn with IASPEI91. This difference is real and caused by the fact that the crust is drastically thicker than IASPEI91 in the source region and under most of the stations causing a delay of almost 5 s in the observed Pn times. Note also the 1-2 s scatter in Pn times relative to an imaginary vertical line. We claim this scatter is due to variations in crustal thickness beneath the Tien Shan. Finally note that on most of the more distant stations Pg is a much larger amplitude phase than Pn. Blunders in catalogs occur for lower signal-to-noise ratio events when Pn becomes invisible and Pg can be incorrectly associated with Pn.

At magnitudes below 4.0 regional distance recordings become the primary source of data for location. As Figure 1 shows, these signals are nearly always emergent and subject to large measurement and earth model errors. When the “constellation” is geometrically inadequate, these errors can be magnified thousands of times relative to a solution that is well constrained by many observations. Geometric problems caused by inadequate constellations can only be solved by adding more seismic stations to record suspect events. The data processing challenge to more effectively utilize the data we do get is to provide calibration data to reduce earth model errors, develop procedures to improve the precision of measurements made on the data, and develop procedures to provide realistic error estimates. The project proposed here addresses all three of these challenges.

Results and Discussion

Catalogue Preparation. The geographic region that stretches from the Middle East to central Asia is an area of monitoring interest. It is also a region in which there have been a number of portable seismic experiments mounted in the last few years due to the interest in the structure and dynamics of the India-Asia collision. We have used data collected in these experiments, as well broadband data collected on a number of national and private networks, along with global seismic stations to produce a high quality regional catalogue with a detection threshold less than magnitude 3. Figure 2 shows the seismic stations that can be integrated, and Table 1 lists details about the various networks.

TABLE 1: Data Source for Catalogue Construction

Network Name	Number of Stations	Dates of Operation
Permanent Global Stations GSN, Geoscope, IMS	14	1995 to the present
Kaznet (network in Kazakhstan)	9	1995 to present, although intermittent
Nanga Parbat, Pakistan (PASSCAL experiment)	10	6/96 thru 9/96
Saudi Portable Network	8	11/95 thru 2/97
Tien-Shan (PASSCAL experiment)	5 stations in Kyrgyzstan 4 stations in China 11 stations in China 18 stations in Kyrgyzstan	9/97 to 8/00 6/98 to 8/00 6/99 to 8/00 7/99 to 8/00

InDepthIII	30 station on the Tibetan Plateau	1994 to 1999
KNET	10 stations in Kyrgyzstan	1995 to present

Each network or experiment listed has been individually processed and has its own catalog. Summing the catalogs from each network or experiment provides over 40,000 events (Figure 3). The fact that high quality seismic stations are located near the seismicity makes the catalogue superior to any other product for the same region in this time period. The final step left to be done is to merge all these independent catalogs into one master catalog at to relocate all events jointly detected by multiple networks using all available phase picks.

Improving location accuracy and error appraisal. A simple summary of this element of our current project is this: we proposed to implement and apply location and error analysis methodologies described in a series of papers by Pavlis and students in the 1980s to the data we proposed to assemble for this region. At the start of the project we had a set of old FORTRAN programs used to validate the original work and a partially finished “generalized earthquake location” (genloc) library of newer code in C. The old FORTRAN code was discarded and we built an entirely new implementation of PMEL (Pavlis and Booker, 1983) and it’s variant SELM (Pavlis and Hokanson, 1985) using the genloc library as a framework. This was a major software development effort as demonstrated by the following table that illustrates scope of this effort.

TABLE 2: Software Development

<i>Program or library name</i>	<i>Brief Description</i>	<i>Number of lines of code</i>
dbpmel	Database driven version of PMEL/SELM	2903
pmelavg	Averages PMEL solutions from multiple grid points	159
cluster	Spatial association program	299
makegclgrid	Builds grid objects for use by cluster	405
makegcl++	C++ version of makegclgrid	139
pmelgrid	Multiprocessor version of pmel	797
db2pfstream	Antelope-based multiprocessor front end	444
pfstream2db	Antelope-based multiprocessor back end	462
libpmel	Shared library for pmel	1570
libgenloc	Generalized Earthquake Location Library	7627
libpfstream	Multiprocessor data flow library	925
libgclgrid	GCLgrid library in C	1139

Central Asia Datasets

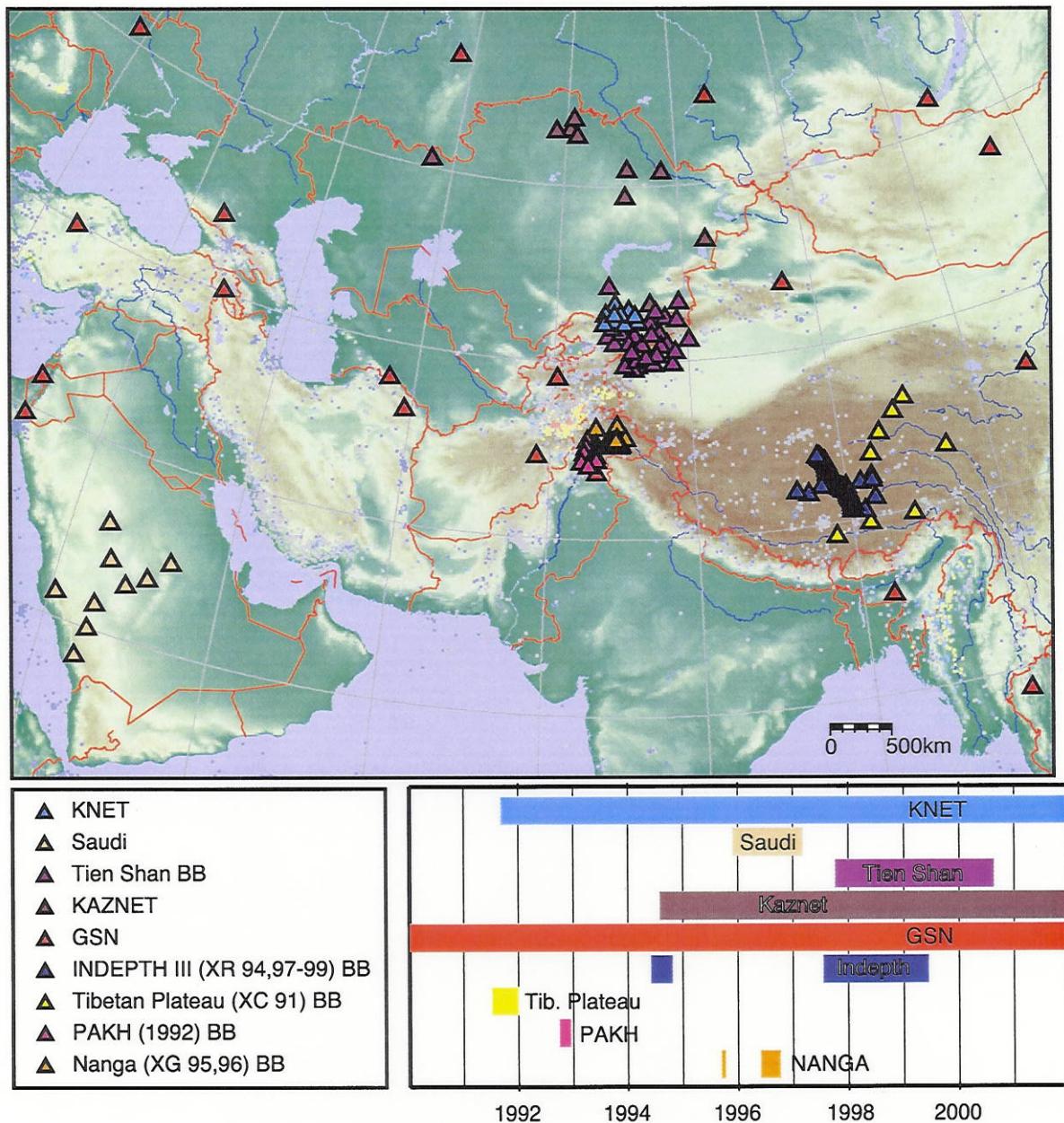


Figure 2. Station map showing all network and portable experiment broadband stations used for preparing the central Asian catalogue. Each network or experiment is designated by different colored triangles. The colored bar graph shows the network or experiment operational times.

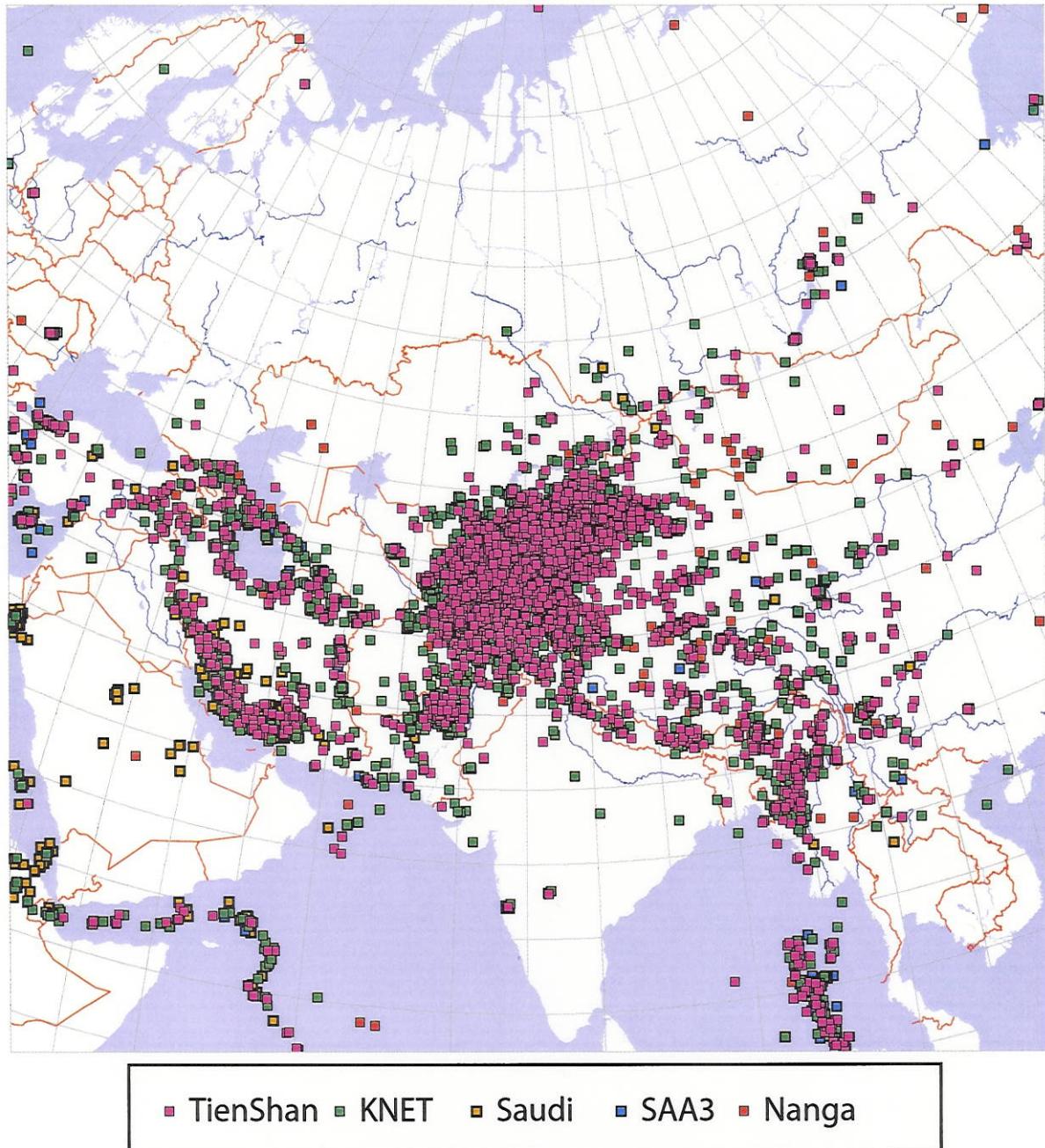


Figure 3. Seismicity recorded by five of the regional networks and portable experiments. Each event is designated by a small colored square corresponding to the source dataset. The squares are not scaled to magnitude.

libgclgrid++	GCLgrid library in C++	2780
	TOTAL	19649

To put this in perspective consider that the listing of this collection of source code is almost 300 pages long has been accomplished in less than three years.

Because these programs define the tools we will use for proposed efforts in relocation it is important to define some details of the implementation. We trust that describing the details in a semi-algorithmic fashion will help clarify how our procedures work and how they differ from other methodologies.

(a) GCLgrid functions. A GCLgrid is a generalized concept for a travel time table in three-dimensions. We developed this concept because the implementation of PMEL/SELM (see below) can utilize 3D velocity models and simultaneously construct empirically derived travel time correction surfaces. Either a travel time table or a set of station-centric corrections relative to a reference model are conveniently expressed as a GCLgrid. Figure 4 is a simple example of such a surface constructed for central Asia. The example shown is a constant depth slice for one station and was constructed with IASPEI91, but it illustrates the concept of a GCLgrid for one station on a uniform grid. (A complete grid is three-dimensional with a set of travel times tabulated for each station in the network.) A standard GCLgrid is a uniform grid warped into spherical geometry. A “standard grid” like that shown is a spherical shell with a translation and rotation of standard geographical coordinates into a more convenient regional reference frame like that shown. We are using this library as a framework for developing a 3D regional-scale travel time calculator. The overall concept is that we will combine this approach with an existing calculator called “ttregions”. The ttregions calculator has a mechanism to divide a region into polygons and to apply a different calculator depending on which region the target point lies in. This will allow us to utilize travel time grids of varying scale and complexity depending on data density.

(b) Spatial Clustering Procedure. Our location methodology utilizes PMEL in a grid mode. The basic idea is that multiple event locations methods like PMEL (Pavlis and Booker, 1983), double-differences (Waldhauser and Ellsworth, 2000), JHD (Douglas, 1967), and Hypocentroidal Decomposition (Jordan and Sverdrup, 1981) all depend upon an assumption that velocity model errors (difference between the real earth and the reference model used for locations) can be approximated as a constant for each station/phase combination (a station correction). This approximation can only be expected to be reasonable when the set of (multiple) events being analyzed with one of these procedures are actually located “close” together. There are outstanding questions about what “close” means quantitatively, but the key issue is that this fundamental assumption is only valid if the velocity model errors

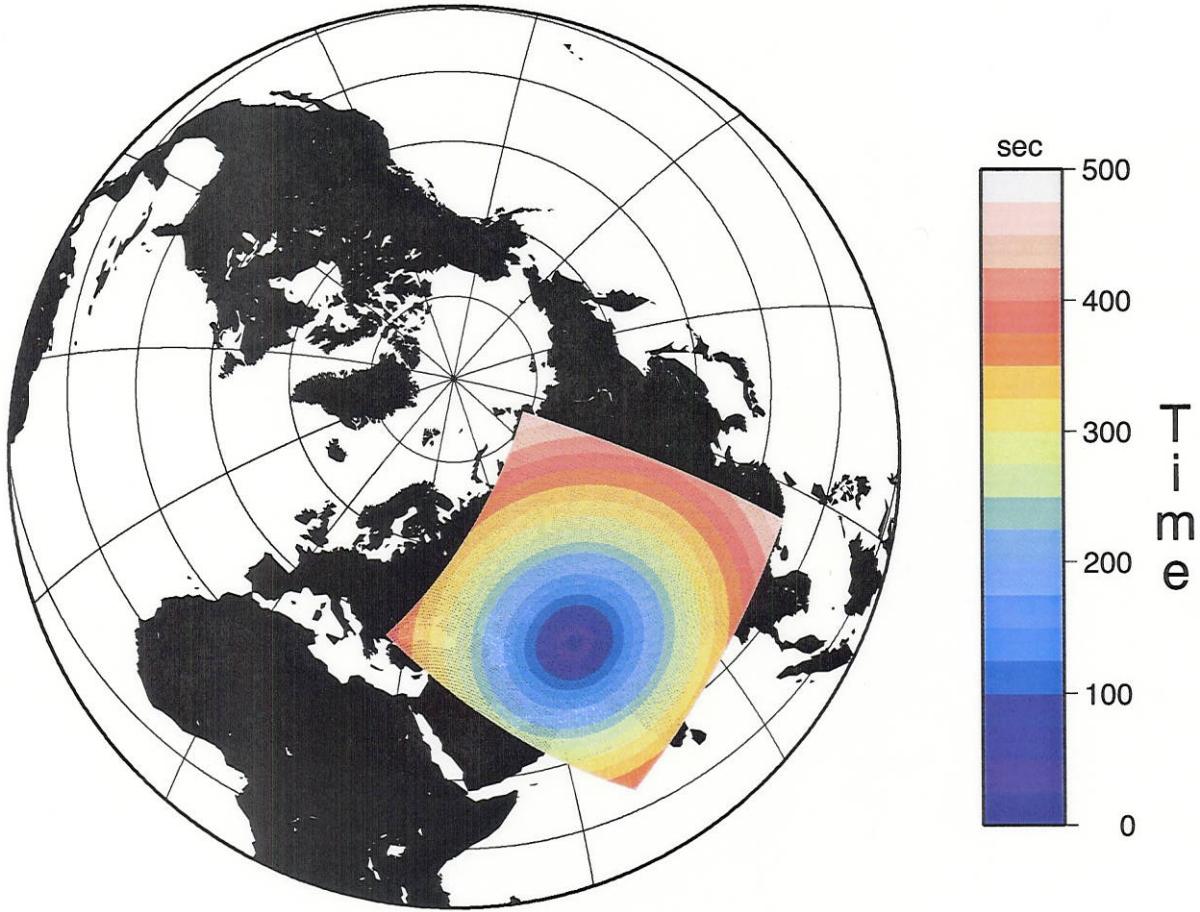


Figure 4. Example of travel time surface defined with a GCLgrid object. This figure illustrates concepts of a GCLgrid by plotting P-wave arrival times predicted by IASPEI91. The standard GCLgrid shown was produced by translating the 0 latitude, 0 longitude point to a point at the center of this grid. Great circles in the generally SW-NE direction are latitude lines and the SE-NW lines are longitude-like lines. This provides an approximation that is as close as possible to an equal-spaced grid for a spherical surface. A full GCLgrid is three-dimensional with sheets on equal depth ellipsoidal surfaces below the reference geoid at $z=0$.

can be characterized as constants and this only happens when the events are within a small distance of each other. The method we use to define close is one based purely on an initial location estimate. The “cluster” program we developed works through a grid of points in 3D (defined by a GCLgrid object) and builds a database table that links each event to one or more grid points. Each such “cluster” then defines an ensemble of events that can be used for one run of PMEL in the traditional sense. Our PMEL implementation, however, is designed to efficiently work through tens of thousands of such ensembles by exploiting the fact that

the problem is “embarrassingly parallel” in the jargon of high performance computing. We exploit this in a parallel processing version of PMEL that utilizes the MPI library (the most common software platform for parallel computing applications). The work for different processors is divided up by grid position with each processor essentially being assigned a list of grid points to process.

(c) Location methodology. Our implementation of PMEL is designed to work through grids of event ensembles. All our applications have used GCLgrids as the locus of grid points for this processing, but any grid of points could, in principle, be used. The program simply loops through all grid points using an exact matrix inversion formula (within the limits of linearization) to solve the multiple event location equations using methods described by Pavlis and Booker (1983). The basic outputs of this procedure are two quantities: (1) revised estimates of the hypocenters of every event in each ensemble, and (2) estimates of path corrections for each station and each seismic phase. Different location estimates occur when an event is associated with multiple grid points. A program called pmelavg averages these multiple estimates in to produce a single “best” solution.

The most important feature of our PMEL implementation is that it provides a complete, exact approach to producing empirically derived 3D travel time curves. A unique feature is the way it implements concepts described in the paper by Pavlis and Hokanson (1985). They showed how to construct a pair of complementary, orthogonal projectors that we will refer to as P_R and P_N . We apply these projectors using the relation

$$\mathbf{s} = P_R \mathbf{s}_{3d} + P_N \mathbf{s}_{data}$$

where \mathbf{s} denotes a vector of path anomalies (one for each station and seismic phase). It has two incarnations here: \mathbf{s}_{3d} is derived from an earth model as the difference between a 1D reference model and a 3D earth model while \mathbf{s}_{data} is the quantity PMEL estimates directly from the data for each grid point. The projectors separate the part of the solution that can be estimated from the data ($P_N \mathbf{s}_{data}$) from that which is fundamentally impossible to determine from the data ($P_R \mathbf{s}_{3d}$). The term $P_R \mathbf{s}_{3d}$ represents the “bias” term that Jordan and Sverdrup (1981) show is impossible to determine from an event cluster. Figure 5 illustrates this concept with a simulation we used to test our PMEL program. A unique feature of our approach is that we use a 3D model only as a reference to correct the bias problem and extract from the data only the information it can possibly give us. Most other approaches being used today do not handle this problem correctly and contain residual biases that are impossible to unravel. Methods based on hypocentroidal decomposition or residual averaging deliberately avoid this problem by using annihilators that remove the dependence on velocity model errors. These methods can only estimate precision relative locations without auxiliary constraints on absolute position. In addition, Wolfe (2002) recently showed that the double-

difference technique of Waldhauser and Ellsworth (2000) uses an implicit spatial averaging scheme that is virtually impossible to unwrap from the results. A promising solution to this problem is the double difference tomography methodology recently introduced by Thurber *et al.* (2002) and Zhang and Thurber (2003). DD tomography is essentially a simultaneous method for solving the same problem we are attacking in pieces. That is, in DD tomography all the data are differenced and the differences are inverted for structure and revised location estimates.

We would argue that our approach has three distinct advantages over DD tomography:

1. Our approach naturally bins the data spatially which simplifies the process of updating a large database of path anomaly estimates as more data is accumulated. This is in contrast to a velocity-model approach which requires simultaneous inversion of an ever expanding dataset to yield increasingly complex models that individually require extensive validation. With our approach as more data accumulates the statistics of the path anomaly estimates improve, but the process of updating the solution is more tractable. Instead of needing to invert an every expanding matrix, we need to invert tens of thousands of smaller matrices that are incremented individually.
2. We separate the bias and “cluster vector” (Jordan and Sverdrup, 1981) components of the solution. This makes it relatively simple to evaluate the effect of using a different model to define the bias term. This is a potentially valuable, empirical way to appraise the scale of bias in locations at different positions inside the earth.
3. Our approach is naturally hierachic and scalable with data density. That is, a first-order feature taught to all earth science students is that earthquakes are highly concentrated at plate boundaries. Anyone’s intuition would be that we should be able to locate earthquakes more precisely in areas where there have been lots of previous earthquakes. Unfortunately, the direction most of the community has taken for improving earthquake locations is to attempt to produce more accurate (which normally means both the velocity estimates and the spatial resolution of the velocity model in 3d) velocity models. This approach leads to a counterintuitive notion that accuracy depends on global coverage and velocity model resolution defined by ray coverage in tomographic inversions. Our approach isolates this problem to its impact on the bias term. By casting the problem as one of estimating station-centric corrections that vary in space, we avoid intermingling of information into a single 3D velocity model. As a result we can, in principle, vary our mesh of control points that define where estimates are made in any way we ultimately choose. In our current work we have used relatively uniform grids of target points, but this is not essential. We have the machinery in place to, for example, estimate corrections on a very dense grids in regions like the Hindu Kush with high levels of

Cluster Simulation Test Results

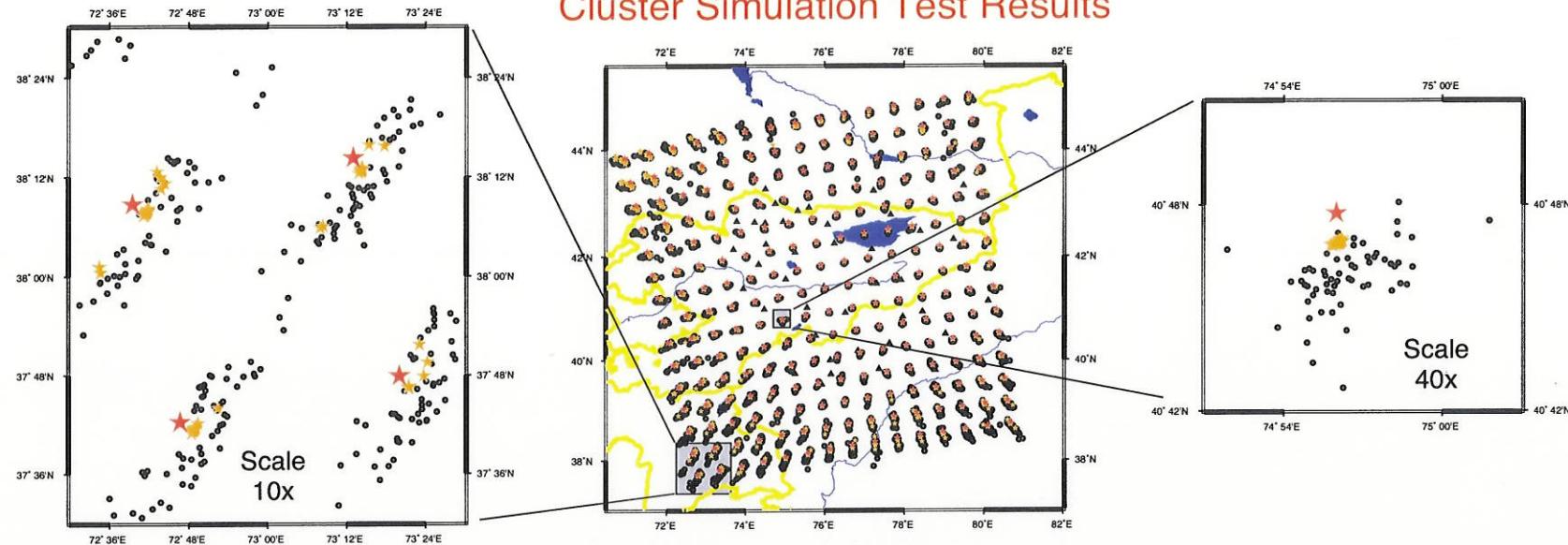


Figure 5. Simulation results illustrating key concepts of our grid-based PMEL implementation. Results shown are for a simulation of events placed at regular intervals defined by a GCLgrid (Figure 4). That is, 75 hypothetical events were defined for each of the points illustrated with the red stars. Travel times were computed to stations in the Tien Shan experimental array (black triangles in the center figure) using IASPEI91 but adding a +1.0 s P offset for stations north of the 42 degree latitude line and a -1.0 s P static offset for stations south of 42 degrees. 25% of the simulated arrivals were randomly discarded. Each of the frames above are map projections of location results on this simulation data set. The center figure is a regional-scale map showing the entire region and the two side figures are blown up map views of results in two geometric configurations. The left figure focuses on points well outside the network of stations with "data". The right figure is for a point surrounded by stations defined in the simulation. In each map the black dots are locations of simulation events computed by standard single event techniques with IASPEI91 but with no path corrections. When PMEL is applied to these data (orange stars), points that are located inside the network (right frame) converge to a tight group that is offset from the true position (red star). This illustrates graphically the cluster ambiguity shared by all multiple event location schemes that attempt to locate events in a cluster. Note that if the true set of station corrections are used as a "reference" (see text) the locations all collapse to a exact point underneath the red star. Note also that for events outside the network (left frame) even after running PMEL some events have large residual errors. This is an inescapable geometric problem. Control on these events is so bad that roundoff errors are enough to significantly distort the solution.

seismicity while estimating a very coarse grid for regions like the Turan Platform with much lower seismicity rates.

Our PMEL implementation also has a number of pragmatic elements designed to allow it to do automated processing in a robust way. That is, as we noted above in our discussion of Figure 1, outliers are a common problem in regional phase measurements. We attempt to handle this problem with two different robust estimation procedures. First, we use residual weighting following the more rigorous concept of M-estimators first introduced into location methodology by Anderson (1978). M-estimators are an effective tool for identifying a small number of outliers in a pool of many measurements, but no outlier detector can easily handle multiple bad data points in relatively small set of measurements. This is, unfortunately, a common problem as it is a defining property of smaller regional events. (For example, you might have one or two very good picks at a close station and 4 or 5 really bad picks at Pn distances.) To handle this we use an F-test to compare the rms residuals of each event against the ensemble average. (We use a corrected rms with the residual of test event removed from the ensemble average.) We discard all data from events with an rms significantly difference from the ensemble average.

Figure 6 demonstrates graphically the unique outputs of our grid-based PMEL. Empirical corrections for P at each station are computed on a semi-regular grid. Note that the gaps are due to lack of data. We cannot compute a correction where there is no data. If we had used a 3D reference model these holes would be filled with corrections computed from the 3D reference model. An added benefit of this approach is shown with residual statistics. We plot variations in rms misfit as a function of space within the network. This demonstrates that we are able to fit the data extremely well inside the network where events are constrained by impulse P and S picks. On the other hand, events outside of the network, which contain large errors due to emergent regional phases, have much larger misfit statistics. Figure 6 shows that the results are not improved as much by including path anomaly corrections and suggest the data from these regions contain large measurement errors.

(d) Waveform correlation procedures. In the past decade a number of studies have demonstrated that waveform cross-correlations methods can significantly improve the precision of relative event locations. (e.g. Got *et al.*, 1994; Shearer, 1997; Fehler *et al.*, 2000; Waldhauser and Ellsworth, 2000; Turber *et al.*, 2001; and Rowe *et al.*, 2002a, 2002b) One should recognize, in fact, that “difference” methods like that of Got *et al.* (1994), Shearer (1997), and Waldhauser and Ellsworth (2000) arose in a natural way out of pair-wise cross-correlation algorithms. That is, the measurement from the correlation of two waveforms from two different events is naturally cast as a precision time-difference measurement. Waveform

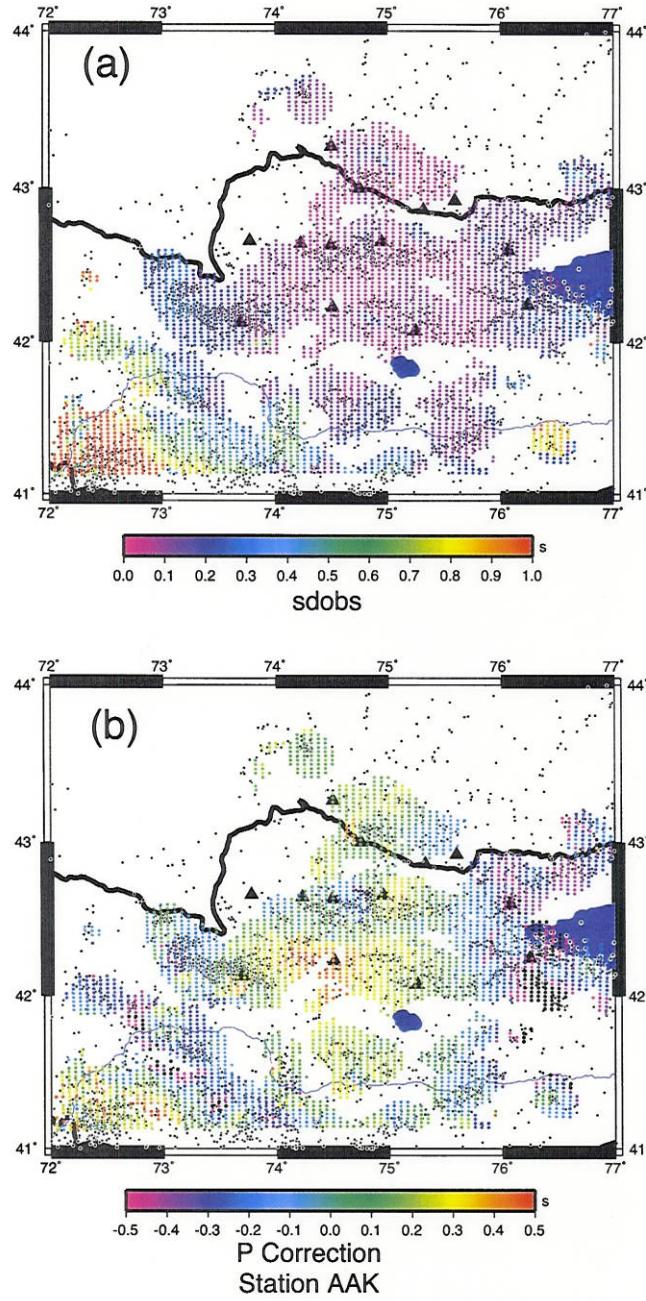


Figure 6. Example of grid-based PMEL processing on KNET. These figures summarize some of the results of applying our grid-based PMEL program to data from the Krghyz Network (KNET). Station locations are shown as black triangles and relocated epicenters are plotted as black dots outlined in white. (a) shows rms residual statistics for each ensemble of events associated with each point in space. The colored dots mark the target position in space and the color of the dot is keyed to rms residuals for that position in space. We see that our solution fits P and S residuals to rms of under 0.1 s for events inside KNET but the misfit grows rapidly for events outside of the network. (b) is a similar plot but color circles are scaled by the size of the P wave correction computed for station AAK. Note that the corrections are readily contourable and define a relatively complex pattern. In both examples white areas are areas with insufficient coverage to warrant a PMEL solution.

correlation, however, should be thought of as two things:

1. Waveform correlation is a measurement tool to improve the precision of arrival time estimates. With waveform correlation subsample measurement precision is theoretically feasible (e.g. Aster and Rowe, 2000) while conventional analyst picking methods are never realistically better than one sample.
2. Waveform correlation is a source array processing problem (Figure 7). The objective is identical to receiver-based array processing; sum the data into a stack that maximizes power or coherence of the stack. What parameterizes the stack, however, is completely different; for a receiver array the primary measurement is a plane-wave slowness vector while for a “source array” the measurement is the relative timing between the signals recorded at the same station.

A critical, limiting issue is that correlation only works when events are close enough and have a source mechanism that is close enough that the recorded signals can actually be correlated. That is, they have to “look alike” or correlation makes no sense because it is literally like comparing apples to oranges.

There is significant reason to believe that waveform correlation may be an exceptionally valuable tool for improving relative location estimates for events constrained mainly by regional phases. Figures 7 and 8 illustrate an example from our recent work on data from the Tien Shan experiment (Figure 1). These events are aftershocks of a magnitude 6.2 event that occurred outside the recording network. Most of the data are characterized by classic, emergent waveforms that are characteristic of regional phases. When correlation methods are applied (Figure 7) we see that the signal are highly correlated and we are able to align the traces wiggle-for-wiggle over a surprisingly long time window for most of the data. Notice that the analyst picks contain large errors and no alignment is seen. The reason, of course, is that the analyst does not see the seismograms assembled in this format, but works on events one at time. Emergent signals are hard to pick consistently because larger events tend to consistently be picked earlier than smaller events (Figure 1). Figure 8 illustrates the impact this has on location estimates. The original catalog forms a diffuse cloud. PMEL applied to the original catalog data helps some. We begin to resolve a fault structure dipping to the northwest that is consistent with the CMT focal mechanism. After correlation many of the events resolve into a much more compact structure at seismogenic depths. An interesting observation is that the results are “better”, as judged by a subjective clustering criteria, when only P wave correlations are used and analyst S picks are used.

We have developed the prototype tools used to produce Figure 7 and by the time the proposed project would start we expect to have an automated tool to do this processing. The program will be driven by the same cluster definitions we used for PMEL. The procedure

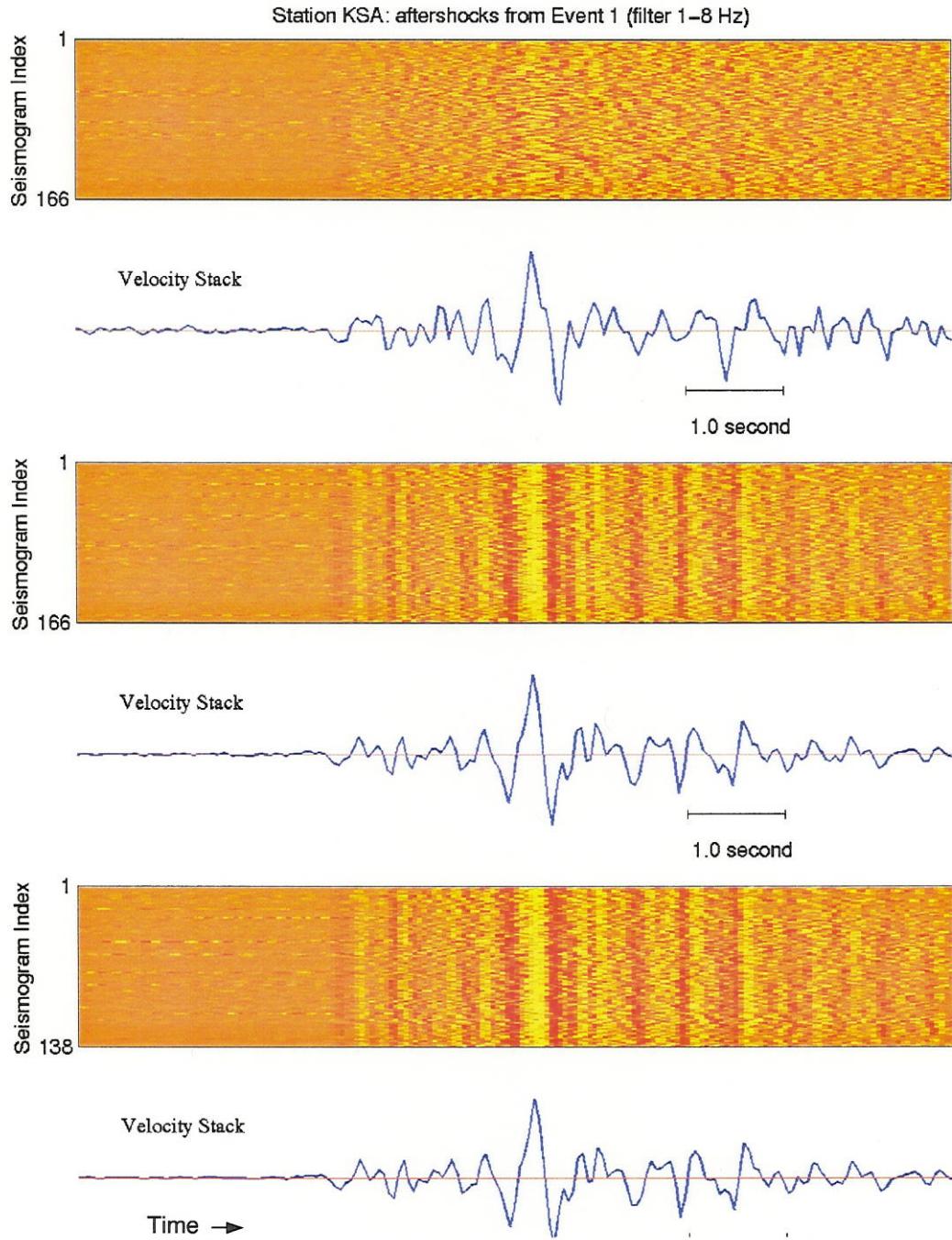
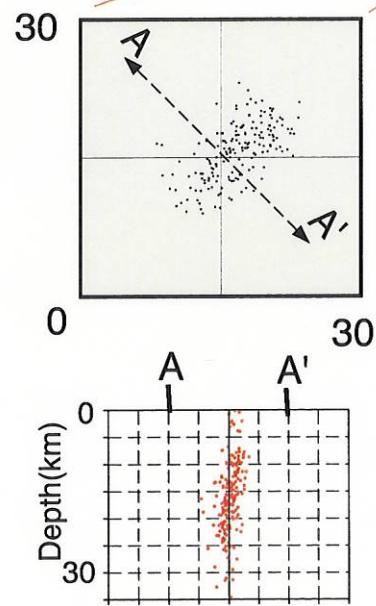
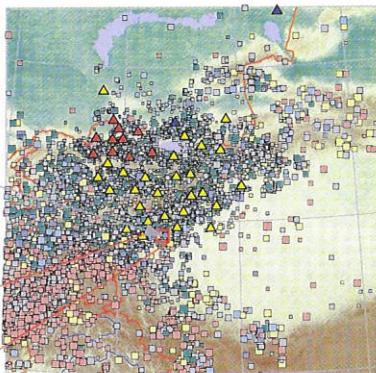
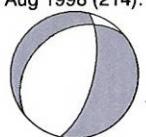
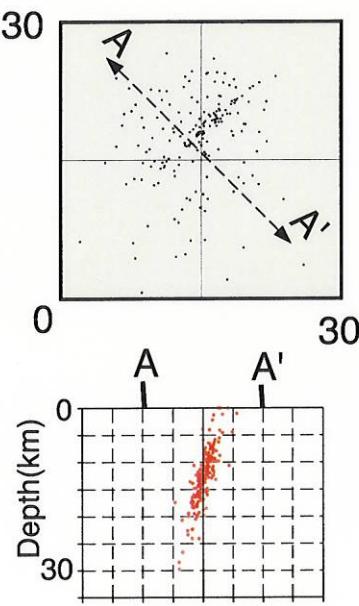


Figure 7. Example of prototype cross-correlation procedure. The colored panel in each part of the figure shows seismograms in an image format with time increasing to the right. Order of the traces is order read. Blue seismograms are the scaled stack of the signals shown in the image window. Top group shows traces aligned using analyst picks. Middle group has traces aligned by largest signal in the ensemble. Bottom uses iterative summed stack as the correlation reference.

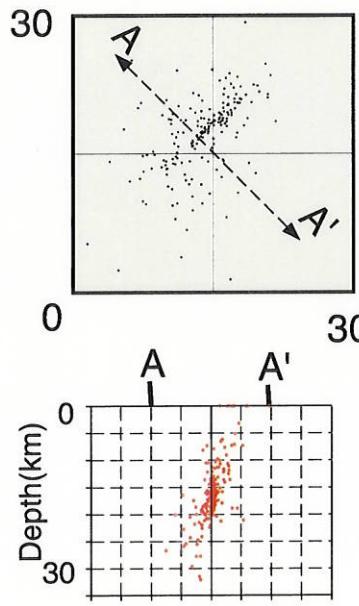
TS M6.1 02 Aug 1998 (214): 231 30 -56



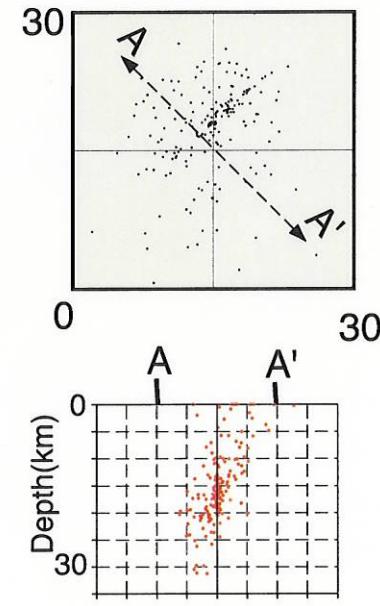
Catalog



PMEL with
analyst picks



PMEL with
P correlations



PMEL with
P&S Correlations

Figure 8. Effects of cross-correlation and PMEL on aftershock locations. We analyzed aftershocks from a sequence following a 6.1 event in the Tarim basin with the CMT mechanism shown. Upper series of figures are map views and the lower set are cross-sections with an approximately 2X vertical exaggeration. Left group were the locations produced interactively by an analyst. Also shown from left to right are: are PMEL locations from the hand picked data, PMEL locations with P wave cross-correlation, and PMEL with P and S wave cross-correlation. In the larger scale base map squares are earthquakes with colors keyed to depth. Triangles are seismic stations: red=KNET, blue=KZNET, and yellow=Tien Shan PASSCAL.

uses a robust stacking algorithm to produce a weighted sum of an ensemble of traces associated with a common grid point. Events are correlated against the weighted stack with the weights defined by an M-estimator and adjusted iteratively until the stack stabilizes.

Conclusions

We have assembled data from a number of portable seismic experiments mounted in the last few years drive largely by an interest in the structure and dynamics of the India-Asia collision. We have used data collected in these experiments, as well broadband data collected on a number of national and private networks, along with global seismic stations to produce a high quality regional catalogue with a detection threshold less than magnitude 3 for most of the middle East and central Asia. Each network or experiment listed has been individually processed and has its own catalog. Summing the catalogs from each network or experiment provides over 40,000 events. We developed a new grid-based implementation of the progressive multiple event location (PMEL). We use a spatial association algorithm to associate each events with one or more grid points within a region. We then apply PMEL to each spatial grouping (cluster) to estimate two quantities: (1) revised estimates of the hypocenters of every event in each ensemble, and (2) estimates of path corrections for each station and each seismic phase. We produce these in hierarchy of scales. We start at the smallest scale to relocate events in the vicinity of each of the individual networks. We use distance weighting to dampen the influence on distant stations that would otherwise skew locations of the largest events that light up a larger area relative to smaller events that are only recorded at the closest stations. We then use the local grid travel times in combination with the “ttregions” travel-time calculator to build a control framework for a regional solution on a coarser grid spanning the Middle-East and most of southern Asia. Events falling inside the local-scale grids will automatically use the local grid corrections as the 3D reference model to derive an improved absolute location framework. This will link the local scale network results into a larger scale framework.

A second parallel effort is to apply waveform correlation methods to the entire dataset. Waveform correlation can dramatically improve measurement precision, but it can only be done successfully when waveforms are similar enough to allow correlation. A key research question is the distance scale length over which waveforms can be correlated. A key practical problem is mixing hand-picked and cross-correlation measurements in the same framework. The working catalogue we produced is the starting point for several critical research questions important for nuclear monitoring.

References

Anderson, K. (1978). Automated processing of local earthquake data, *PhD dissertation*, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Aster, R. and C. Rowe (2000). Automatic phase refinement and similar event association in large seismic datasets, in *Advances in Seismic Event Location*, ed. C.H. Thurber and N. Rabinowitz, Kluwer, Netherlands, 231-263.

Bergmann and Engdahl(2002). Probability density functions for secondary seismic phase arrivals, in *24th Seismic Research Review*, p252-260.

Douglas, A. (1967). Joint epicentre location, *Nature*, 215, 45-48.,

Fehler, M., W. W. Phillips, L. House, R. H. Jones, R. Aster, and C. Rowe (2000). Improved relative locations of clustered earthquakes using constrained multiple event location, *Bull. Seism. Soc. Amer.*, 90, 775-780.

Got, J. L., J. Frechet, and F.W. Klein (1994). Deep fault geometry inferred from multiplet relative location beneath the south flank of Kilauea, *J. Geophys. Res.*, 99, 15375-15386.

Jordan, T. H. and K. A. Sverdrup (1981). Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seism. Soc. Amer.*, 71, 1105-1130.

Pavlis, G. L. (1992). Appraising relative earthquake location errors, *Bull. Seism. Soc. Amer.*, 82, 836-859.

Pavlis, G. L. (1986). Appraising earthquake hypocenter location errors: a complete, practical approach for single event locations, *Bull. Seis. Soc. Amer.*, 76, 1699-1717.

Pavlis, G. L. and N. B. Hokanson (1985). Separated earthquake location, *J. Geophys. Res.*, 90, 12,777-12,789.

Pavlis, G. L. and J. R. Booker (1983). Progressive multiple event location (PMEL), *Bull. Seis. Soc. Amer.*, 73, 1753-1777.

Rowe, C. A., R. C. Aster, B. Borchers, and C. J. Young (2002a). An automatic, adaptive algorithm for refining phase picks in large seismic data sets, *Bull. Seism. Soc. Amer.*, 92, 1660-1674.

Rowe, C. A., R. C. Aster, W.S. Phillips, R. H. Jones, B. Borchers, and M.C. Fehler (2002b). Using automated, high-precision repicking to improve delineation of microseismic structures at the Soultz geothermal reservoir, *Pure Appl. Geophys.*, 159, 563-596.

Shearer, P. M. (1997). Improving local earthquake locations using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence, *J. Geophys. Res.*, 102, 8269-8283.

Thurber, C., C. Trabant, F. Haslinger, and R. Hartog (2001). Nuclear explosion locations at Balapan, Kazakhstan, nuclear test site: the effects of high-precision arrival times and three-dimensional structure, *Phys. Earth Planet. Int.*, 123, 283-301.

Thurber, C. H., H. Zhang, C. A. Rowe, and W. J. Lutter (2002). Methods for improving seismic event location processing, *Proceedings of the 24th Seismic Research Review – Nuclear Explosion Monitoring: Innovation and Integration, Vol I, Rept. LA-UR-02-5048*, 438-447.

Waldhauser, F. and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Methds and application to the Northern Hayward Fault, California, *Bull. Seism. Soc. Amer.*, 90, 1353-1368.

Wolfe, C. J. (2002). On the mathematics of using difference operators to locate earthquakes, *Bull. Seism. Soc. Amer.*, *in press*.

Zhang, H. and C. Thurber (2003). Double-difference tomography: Methods and application to the Hayward Fault, California, *Bull. Seism. Soc. Amer.*, *submitted*.

DISTRIBUTION LIST
DTRA-TR-12-63

DEPARTMENT OF DEFENSE

DEFENSE TECHNICAL
INFORMATION CENTER
8725 JOHN J. KINGMAN ROAD,
SUITE 0944
FT. BELVOIR, VA 22060-6201
ATTN: DTIC/OCA

**DEPARTMENT OF DEFENSE
CONTRACTORS**

EXELIS, INC.
1680 TEXAS STREET, SE
KIRTLAND AFB, NM 87117-5669
ATTN: DTRIAC